IS COLLUSION-PROOF PROCUREMENT EXPENSIVE?*

GAURAB ARYAL† AND MARIA F. GABRIELLI‡

ABSTRACT. Collusion among bidders adversely affects procurement cost and in some cases efficiency, and it seems collusion is more prevalent that we would like. Statistical methods of detecting collusion just using bid data, in a hope to deter future collusion, is perilous, and access to additional data is rare and often always after the fact. In this paper, we estimate the extra cost of implementing a new procurement rule proposed by Chen and Micali [2012] that is robust to collusion and always guarantees the efficient outcome. The rule requires bidders to report their coalition and to ensure incentive-compatibility, the mechanism allows them to attain rents. We estimate this rent using data from California highway construction and find it to be anywhere between 1.6% to 5%. Even after we factor in the marginal excess burden of taxes needed to finance these rents, the cost ranges between 2.08% and 6.5%, suggesting that there is a room to think about running this new auction, suggesting we should consider this auction.

Keywords: Collusion; Procurements; Collusion-Proof Auction; Local Polynomial Estimator.

JEL: C1, C4, C7, D44, L4.

1. INTRODUCTION

Auction is a widely used method, to buy and sell goods and services, in both private and government sectors. Some of these markets (timbers auctions, highway constructions, off-shore wild cat auctions, etc.) have been studied extensively in empirical auction; see Athey and Haile [2006] and the references therein. Whether or not auction achieves its goal of maximum revenue (or minimum cost) and efficiency, depends crucially on the competition among bidders, Ausubel and Milgrom [2006]. It is, however, well known that auction is susceptible to collusion, Marshall and Marx [2009],

*We thank Jing Chen and Paulo Somaini for helpful comments. Valentina Reig provided an excellent research assistance. All errors are our own.

† University of Virginia. e-mail: aryalg@virginia.edu

‡ CONICET and Universidad Nacional de Cuyo. e-mail: florgabrielli@gmail.com.

and therefore should be factored in while choosing auction rules. In some cases, like the recent Libor scandal, the extent of the damages inflicted on public by collusion has been very large and has also attracted attention from press, see Taibbi [2012, 2013]. There are papers that have studied the effect of collusion on revenue or cost, some of them include Comanor and Schankerman [1976]; Feinstein, Block, and Nold [1985]; Lang and Rosenthal [1991]; Porter and Zona [1993]; Baldwin, Marshall, and Richard [1997]; Bajari [2001]; Porter and Zona [1999]; Pesendorfer [2000]; Bajari and Ye [2003]; Marshall and Meurer [2004]; Asker [2008]; Harrington [2008]; Marshall and Marx [2012]. Even though it is impossible to quantify the likelihood of collusion, there seems to be an overwhelming evidence to suggest that the likelihood is non negligible. Therefore collusion should be of first-order importance in any empirical auction paper whose aim is to use the data to estimate the parameters that can then be used to design optimal auctions.

One of the main difficulties when it comes to collusion is that collusive rings tend to be secretive and sly, and detecting their presence from only bids data is perilous, Aryal and Gabrielli [2013]. Moreover, it is rare for researchers to get access to any other information pertaining to the cartel like those in [see Porter and Zona, 1999; Asker, 2008; Conley and Decarolis, forthcoming]. These are, however, 'after-the-fact' studies that do not help the auctioneer to deter collusion. As seen with the Libor scandal, Hou and Skeie [2013], even when the identity of the suspected colluders and the damages are known, proving any wrongdoing is hard and often a drawn-out and expensive process. This begs the question as to why don't the auctioneer use collusion-proof auction. Green and Laffont [1979]; Laffont and Martimort [1997, 1998, 2000] have shown that it is very hard to find such auction.

Keeping these difficulties in mind, in this paper we want to suggest that we should use a new auction rule proposed by Chen and Micali [2012] (henceforth, C-M). C-M propose a variation of Vickrey auction which requires firms to report their coalition and to ensure incentive-compatibility, the mechanism allows firms to attain rents. This modified Vickrey auction is resilient to collusion, induces efficient allocation, in dominant strategies while maintaining the ex post individual rationality for all parties. Using

California highway procurements data, we estimate this rent to be anywhere between 1.6% to 5%. This is the price we have to pay to ensure efficiency in auction even when bidders collude in an arbitrary way. Even if we factor in the marginal excess burden of taxes needed to pay for these rents, the increase in cost is between 2.08% and 6.5%.

To calculate the rent, we first use the data from California Highway procurements to estimate the bidders' cost. For that we estimate the bid distribution and bid density using the first step of Aryal, Gabrielli, and Vuong [2015] who propose a $\sqrt{n}-$ consistent semiparametric estimation method. They propose using local polynomial estimation method (henceforth, LPE) of Fan and Gijbels [1996], instead of the widely used kernel-smoothed density estimators, because unlike the kernel density estimator LPE estimators are consistent even at the boundary and therefore do not require trimming. Then following Guerre, Perrigne, and Vuong [2000] we use a first-price procurmenet auction model with asymmetric bidders to identify and estimate the cost parameter for each construction company. Since it is possible that the data itself have collusive bidders, we have to control for that in our estimation. For that we use the tests in Bajari and Ye [2003] to determine a set of potential colluders by finding those whose bidding pattern depart from competitive bidding. Then, using the estimated costs, in a counterfactual exercise we implement C-M auction and calculate the change in total cost. Out of 413 bidders, we flag 15 who fail one of the two competitive-bidding (exchangeability and independence) tests. The worse case for Caltrans is if all of these bidders form one single bidding ring and hence we treat all of these 15 bidders as members of one ring. We also recognize that forming and maintaining a large bidding ring is difficult, so we explore other ways to restrict the size of the ring. Using the idea that ring members tend to bid simultaneously to provide cover bids, we find that only 4 out of the above 15 fail the test.

Depending on whether there are only 4 or 15 bidders in a bidding ring, we find that using the new auction will increase Caltran's cost by respectively, (approximately) $34 millions to $107 millions, i.e., 1.6% to 5% of the total construction expenditure in our sample. If, however, we take into account that to pay for the additional expenses the (state or federal) government

would have to collect taxes, which leads to deadweight loss, the previous numbers increase. While estimates of the deadweight loss due to taxes are hard to come by, Ballard, Shoven, and Whalley [1985] estimate the marginal deadweight loss to be a sizable 0.3. This means if Caltrans were to implement the new auction it would increase the (social) cost by at least 2.08% bounded above by 6.5%.

The main message of this paper is: a) collusion is bad for the economy, especially when it comes to public procurements; b) collusion seems to be very prevalent, and even if we do not have a precise idea about how frequent it is, partly because it depends on the nature of the product or services and partly because most collusion goes undetected, casual observation suggests that it is quite frequent and should be an important feature in optimal auction design; c) auctions that are robust to collusion are hard to design, Chen and Micali [2012] is one such paper that proposes a new auction rule, but it requires the auctioneer to provide some rent to the colluders to ensure that they tell the truth about their membership in any collusive rings; e) we estimate this extra cost (rent) required to run this new auction using data from California highway construction and estimate it to be no more than 6.5% and no less than 2.08%, after accounting for any distortion resulting from higher taxes required to pay the rents, which seems to be reasonable. We should not reject C-M auction outright just because the cost will increase by 6.5%, because the right comparison should be between the C-M auction and losses due to collusion, that we cannot even estimate, and not between the C-M auction and competitive bidding. If bidders bid competitively, C-M auction would be the same as second-price auction.

Having said that, we understand the legal and philosophical difficulty that might arise in giving rents to those who collude and committing to never using this information in prosecution. Such a leniency program, however, is not without precedence. The Leniency Program run by the Antitrust Division of the Department of Justice is one such example, and it is often touted as one of the most important investigative tool to deal with cartel activities.[1] The program has been successful even outside the US; see Borrell, Jiménez, and Garcia [2013]. When faced with the possibility that bidders

[1] See http://www.justice.gov/atr/public/criminal/leniency.html.

could collude without we ever knowing it, and with possibly unrestricted and often unknown damages, it looks reasonable to say that if providing some rent to potential colluders ensures that we always get efficient outcome then may be we should. But to be able to shed light on how one can extend these leniency program to include C-M auction needs further probing and research, and is left for future.

This paper is organized as follows: Section 2 outlines the models, identification and estimation; Section 3 introduces C-M auction; Section 4 explains the data; Section 5 presents and discusses the empirical findings. Choice of bandwidths, tables and figure are in the Appendix A-1.

2. Model, Identification and Estimation

In this section we consider a procurement auction, i.e. a low-price sealed bid auction, with asymmetric (regular and fringe) bidders. The section is divided into two subsections. The first subsection introduces the model and identification, and the second subsection deals with estimation.

2.1. **Model and Identification.** For every auction $\ell = 1, 2, \ldots, L$, a single and indivisible project is procured to $N_\ell \geq 2$ risk neutral bidders using first price low bids mechanism. The essential characteristic of the project for each auction is summarized by a random variable $X_\ell \in \mathbb{R}_+$, which for us will be the engineer's estimate of the project. We assume that there are two types of bidders, $k = 0, 1$, with $n_{k\ell}$ type k bidders in $\ell-$auction, such that $N_\ell = n_{0\ell} + n_{1\ell}$. In every auction ℓ a type k bidder draws his/her cost, i.i.d. across all other bidders, from $F_k(\cdot|X_\ell, N_\ell)$. The set of observables W is

$$W := \left\{X_\ell, n_{0\ell}, n_{1\ell}, \{b_{0i}\}_{i=1}^{n_{0\ell}}, \{b_{1i}\}_{i=1}^{n_{1\ell}}, \right\}, \ell = 1, 2, \ldots L.$$

where b_{ki} is the bid by type $k \in \{0, 1\}$ bidder $i \in n_{k\ell}$. We assume that:

Assumption 1. *(A1)*

(1) *Exogenous Participation:* $F_k(\cdot|X, N) = F_k(\cdot|X)$ *for* $k = 0, 1$.
(2) *For each* ℓ *and each* $k \in \{0, 1\}$ *the variables* $C_{ki\ell}, i \in n_{k\ell} \sim$ *iid* $F_k(\cdot|\cdot)$ *with density* $f_k(\cdot|\cdot)$ *conditional on* X_ℓ.
(3) *An auction* ℓ *has* $N_\ell \in \{\underline{n}, \overline{n}\}$ *risk-neutral bidders with* $\underline{n} \geq 2$.

(4) *The three- dimensional vector* $(X_\ell, n_{0\ell}, n_{1\ell}) \sim iid\ Q_m(\cdot,\cdot,\cdot)$ *with density* $q_m(\cdot,\cdot,\cdot)$ *for all* $\ell = 1,2,\ldots L$*, where* $Q(\cdot)$ *is a product of absolutely continuous measure and a counting measure.*

(5) *The bids* $B_k \sim iid\ G_k(\cdot|X_\ell, N_\ell)$ *with density* $g_k(\cdot|X_\ell, N_\ell) > 0, k \in \{0,1\}$.

A strategy for bidder i of type k is a strictly increasing, type symmetric bidding strategy $s_k : [\underline{c}, \overline{c}] \to [\underline{b}, \overline{b}]$. From Guerre, Perrigne, and Vuong [2000] it follows that the type k bidder i solves

$$\begin{aligned}\max_b \Pi_k(b_i; c_i, X_\ell, N_\ell) &= \max_{b_i}(b_i - c_i) \prod_{j\in n_k\setminus\{i\}} (1 - F_k(s_k^{-1}(b_i)|X_\ell))^j \prod_{j\in n_{k'}} (1 - F_{k'}(s_{k'}^{-1}(b)|X_\ell))^j\\ &= \max_{b_i}(b_i - c_i) \prod_{j\in n_k\setminus\{i\}} (1 - G_k(b_i|X_\ell, N_\ell))^j \prod_{j\in n_{k'}} (1 - G_{k'}(b_i|X_\ell, N_\ell))^j,\end{aligned}$$

where $k \neq k' \in \{0,1\}$ and $G_k(b|X_\ell, N_\ell) = F_k(s_k^{-1}(b)|X_\ell)$ is the probability that bidder $j \in n_{k\ell}\setminus\{i\}$ will bid less than b, and likewise for k'. The first order condition for $i \in n_{k\ell}$ is

$$(b_{ki} - c_{ki}) = \frac{1}{(n_{k\ell} - 1)\frac{g_k(b_{ki}|X_\ell,N_\ell)}{1-G_k(b_{ki}|X_\ell,N_\ell)} + n_{k'}\frac{g_{k'}(b_{ki}|X_\ell,N_\ell)}{1-G_{k'}(b_{ki}|X_\ell,N_\ell)}}. \tag{1}$$

This first order condition with the boundary conditions $s_k(\overline{c}) = \overline{b}, k = 0,1$ uniquely characterizes optimal bidding strategy for all bidders. The model structure is the type specific conditional distribution of cost $\{F_k(\cdot|X_\ell)\}$ for $k = 0,1$ given X. But since the data provide information on the characteristics of the project that is being procured, X_ℓ in the ℓ^{th} project, we can consider only the type specific conditional cost distribution $F_k(\cdot|X_\ell), k = 0,1$ as the structural parameter. The question of identification is to ask if there are two pairs of cost distributions $\{F_0(\cdot|X_\ell), F_1(\cdot|X_\ell)\}$ and $\{F_0'(\cdot|X_\ell), F_1'(\cdot|X_\ell)\}$ that are observationally equivalent.

Evaluating (1) at the estimated bid distribution and densities, we see that for each auction ℓ, bid b_{ki} uniquely determines the cost

$$c_{ki} = b_{ki}^\ell - \frac{1}{(n_{k\ell} - 1)\frac{g_k(b_{ki}^\ell|X_\ell,N_\ell)}{1-G_k(b_{ki}^\ell|X_\ell,N_\ell)} + n_{k'\ell}\frac{g_{k'}(b_{ki}^\ell|X_\ell,N_\ell)}{1-G_{k'}(b_{ki}^\ell|X_\ell,N_\ell)}}, \tag{2}$$

thereby identifying $\{F_0(\cdot|X_\ell), F_1(\cdot|X_\ell)\}$ that is consistent with the data.

2.2. **Estimation.** In the first step we follow Aryal, Gabrielli, and Vuong [2015] and estimate the conditional bid distributions $G_k(\cdot|X, N)$ and the bid

densities $g_k(\cdot|X,N)$ given the engineer's estimate X and the set of bidders N, using Local Polynomial Estimation (LPE) method. We begin by introducing the LPE, see Fan and Gijbels [1996] for more.

Consider a bivariate i.i.d. data $\{X_i, Y_i\}_{i=1}^n$. Our interest is the regression function $m(x_0)$ and its derivatives $m'(x_0), m''(x_0)$ and so on till $m^p(x_0)$. Hence, we regard the model $E[Y|X] = m(X)$. Under the assumption that $(p+1)^{th}$ derivative of $m(\cdot)$ exists at $x = x_0$, LPE can approximate $m(\cdot)$ by a polynomial of order p. Taylor expansion gives

$$m(x) \approx \sum_{j=0}^{p} m^j(x_0)\frac{(x-x_0)^j}{j!},$$

and this polynomial is fitted locally by a weighted least squares regression that minimizes

$$\sum_{i=1}^{n}\{Y_i - \sum_{j=0}^{p}\beta_j(x-x_0)^j\}^2 K_h(X_i - x_0),$$

where h is the bandwidth, $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)/h$ with K a kernel function. If $\hat{\beta}_j, j = 0,\ldots,p$ is the solution to the weighted least squares then $j!\beta_j(x_0)$ is the estimator for $m^j(x_0), j = 0,\ldots,p$. The exact form used for our estimation is given in Appendix (A-1). We make the following assumptions for estimation.

Assumption A3:

(i) *The kernels $K_G(\cdot)$, $K_{0g}(\cdot)$ and $K_{1g}(\cdot)$ are symmetric with bounded hypercube supports and twice continuous bounded derivatives with respect to their arguments,*

(ii) $\int K_G(x)dx = 1$, $\int K_{0g}(x)dx = 1$, $\int K_{1g}(b)db = 1$

(iii) *$K_G(\cdot)$, $K_{0g}(\cdot)$ and $K_{1g}(\cdot)$ are of order $R-1$. Thus moments of order strictly smaller than $R-1$ vanish.*

Assumption A4: *The bandwidths h_G, h_{1g} and h_{2g} satisfy*

(i) $h_G \to 0$ *and* $\dfrac{Lh_G^d}{\log L} \to \infty$, *as* $L \to \infty$,

(ii) $h_{0g} \to 0$, $h_{1g} \to 0$ *and* $\dfrac{Lh_{0g}^d h_{1g}}{\log L} \to \infty$, *as* $L \to \infty$.

From this assumption it is clear that it is possible to choose the optimal bandwidths, i.e. the bandwidths proposed in Stone [1982]. As mentioned in the introduction, an important advantage of using LPE over Kernel density estimator as in Guerre, Perrigne, and Vuong [2000] is that LPE is consistent even at the boundary of the support. Kernel density estimators are inconsistent at the boundaries and require some form of boundary correction. One widely used method is trimming, for which we need to specify a "boundary bandwidth," which requires the knowledge of the location of the boundaries.[2] None of that is necessary for us, because of the LPE.

We have 3 conditioning variables, one that is continuous and two that are discrete. Thus, we have to adapt the definition of the LPE to the present case. However, the discrete variables do not affect the asymptotic properties of the estimator, [see Bierens, 1987; Guerre, Perrigne, and Vuong, 2000, footnote 12], so in order to choose the optimal bandwidth the relevant number of covariates to consider is the number of continuous variables. Similar observation is made by Abadie and Imbens [2006]. Let p be the number of continuous variables and d be the total number of conditioning variables. For our application, $d = 3$ and $p = 1$. Let $\hat{\psi} = \hat{g}(\cdot|\cdot,\cdot)/1 - \hat{G}(\cdot|\cdot,\cdot)$ be the estimator of $\psi(\cdot|\cdot,\cdot,\cdot) = g_k(\cdot|\cdot,\cdot,\cdot)/1 - G_k(\cdot|\cdot,\cdot,\cdot)$. From Proposition 1 in Guerre, Perrigne, and Vuong [2000] we know that $G_k(\cdot|\cdot)$ is $R+1$ times continuously differentiable on its entire support and therefore $g_k(\cdot|\cdot)$ is R times continuously differentiable on its entire support as well.[3] Given the smoothness of each function we propose to use a LPE (R), i.e. a LPE of degree R, for $G_k(\cdot|\cdot,\cdot,\cdot)$ and a LPE $(R-1)$ for $g_k(\cdot|\cdot,\cdot,\cdot)$. Aryal, Gabrielli, and Vuong [2015] have shown that the bid distribution is consistent and following Guerre, Perrigne, and Vuong [2000] it is easy to see that the estimated costs are strongly consistent. The exact econometrics model and the selection of optimal bandwidths are explained in Appendix A-1.

[2] Other correction methods include: the reflection method Cline and Hart [1991]; Schuster [1985]; Silverman [1986], the boundary kernel method Gasser and Müller [1979]; Gasser, Müller, and Mammitzsch [1985]; Jones [1993]; Zhang and Karunamuni [2000].

[3] From Proposition 1 by Guerre, Perrigne, and Vuong [2000], we also know that the conditional density $g_0(\cdot|\cdot,\cdot)$ is $R+1$ times continuously differentiable on a closed subset of the interior of the support and thus the degree of smoothness closed to the boundaries and at the boundaries of the support is not $R+1$.

3. Collusion Proof Auction

In this section we introduce the C-M auction. We present only the main (and necessary) result(s) from the paper; anyone interested in formal treatment of the topic should consult C-M. Let there be $N < \infty$ risk-neutral bidders in an independent private value, low bid auction. Each bidder draws i.i.d cost $C \sim F(\cdot)$.[4] Let $\mathbb{C}$ represent the partition of players such that each element of the partition represents a coalition such that every singleton $\{i\} \in \mathbb{C}$ is an independent bidder, and M is a generic element. Here in the example $\mathbb{C} = \{\{1,2\},\{3\}\}$. Let $\mathcal{M} = \{N, F(\cdot), \mathbb{C}\}$ be the context of the game and is commonly known by all bidders. Moreover, we assume that for every coalition $M \in \mathbb{C}$, the $|M|-$ tuple cost profile $C_M = \{C_i : i \in M\}$ is common knowledge only amongst the bidders in that coalition. The seller, however, only knows $\{N, F(\cdot)\}$.

Let $\{A_i, P_i\}_{i=1}^{N}$ be an allocation and pricing rule, where $A_i \in \{0,1\}$ such that $\sum_{i\in N} A_i = 1$ and P_i is the price paid by bidder i. We assume that all coalitions are efficient and hence when the ex-post utility of a bidder i is $(P_i - C_i)A_i$, the utility of the coalition M is the sum across the members, i.e. $u_M = \sum_{i\in M}(P_i - C_i)A_i$. Each member $i \in M$ acts to maximize $u_M(\cdot)$.

Definition 1. *An auction, for a context $\mathcal{M}$, is directly collusive if the set of pure strategies for $i, s_i(\cdot)$ consist of the set of all mapping from $C \mapsto (C, M)$.*

So, a bidder with cost C reports his cost and the coalition M. Let $u_M(s)$ denote the total utility of coalition M when everyone uses symmetric bidding strategy $s(\cdot)$. Now, we are in a position to define dominant-strategy truthfulness and coalitional rationality.

Definition 2. *An auction is collusively dominant-strategy truthful if, for all coalition $M \in \mathbb{C}$ and all strategy profiles $s_M := \{s_i(\cdot) : i \in M\}$ and $s_{-M} := \{s_j(\cdot) : j \notin \mathbb{C}\backslash\{M\}\}$, $\forall i \in M : u_M((C_i, M), s_{-M}) \geq u_M((C_i', M'), s_{-M})$. and is coalitionally rational if $u_M((C_i, M), s_{-M}) \geq 0$.*

Let $s(\cdot) = \{(C_1, M_1), \ldots, (C_N, M_N)\}$ be an action profile. Then a disagreement (in $s(\cdot)$) is an ordered pair (i, j) such that $M_i \ni j$ but $M_j \not\ni i$. In other

[4] For notational ease, we treat all bidders to be symmetric, extending it to asymmetric bidders is straightforward.

words, we say that (i, j) is disagreement if i claims to be a part of collusion ring M_i that contains j but j does not reciprocate. Given a profile $s(\cdot)$ the outcome (A, P) is computed as follows: First, there is the punishing phase if there is any disagreement, in which case $A_i = 0$ for all i. Then to determine the price we start with $P_i = 0$ and for each disagreement (i, j) charge $P_i = P_i + 2t$ and $P_j = P_j - t$, while keeping t with the seller. (Since punishment is off-the-equilibrium path it does not affect the estimation results.) Second, when there is no disagreement we initiate the standard phase where from the reported coalitions the coalition partition $\mathbb{C}$ is constructed. Then, the lowest bidder wins the auction and we determine the winning coalition M^* and charge

$$P_i = \begin{cases} 0 & \text{if } A_{M^*} = \sum_{j \in M^*} A_j = 0 \\ C_{1:(N \backslash M^*)} & o/w \end{cases}$$

to every bidder i, where $C_{1:(N\backslash M^*)}$ is the lowest bid from the bidder $j \notin M$.

Theorem 1. *C-M: The mechanism outlined above is (a) Collusive dominant-strategy truthful; (b) Coalitional rational and ; (c) Efficient.*

4. Data

Below we describe our data and the implementation of Bajari and Ye [2003] tests to determine bidders whose bidding pattern is at odds with what we would expect under competitive bidding.

Data Source. The data consist of the Highway procurements in the state of California between January 2002 and January 2008. The rights to maintain and construct highways and roads are granted through sealed low-bid auctions (procurements) by the California Department of Transportation (Caltrans). The data is available from Caltrans official-webpage, and contains information about the characteristics of the projects, the name of bidders and their submitted bid.

Timing. First, during the *advertising period* that can last between three to ten weeks, depending on the size of the project, the Caltrans Headquarters Office Engineer announces a project and solicits bids. A project can be: asphalt repaving, road paving, bridge reconstruction, striping the highway,

constructing, replacing and widening brides, storm damage repair, etc. Potential bidders express their interest by buying the project catalogue – only those with a catalogue can bid. Second, sealed bids are received only from among the potential bidders. Third, on the *letting day*, bids are ranked and the project is awarded to the lowest bidder, provided that the bidder fulfills other criteria determined by either the federal or state laws or both.

Bidders and Projects. Bidders can be asymmetric in terms of their frequency of participation across different projects. This could be because (on average) some bidders are more efficient (have lower cost) than the others. We assume that bidders can be divided into two types: the fringe (type $k = 0$) bidders and the regular (type $k = 1$) bidders. So a fringe (main) bidder is someone who participates less (more) frequently and in smaller (bigger) projects. There are 823 bidders who bid at least once. Following the literature, we define regular bidders as bidders who have a nontrivial (at least 1%) revenue share in the market and participate frequently (at least 3%). Twenty-one bidders satisfy this criteria and will henceforth be called the type 1 bidders while the remaining are type 0.

There are in total 2,152 unique projects awarded by Caltrans, for a total of $7.645 billions, but some auctions have only one bidder. Once we drop those single bidder auctions we are left with 1,907 projects. For every project we know the county of the site of job, and because we know bidders' main offices we can calculate the distance (in miles) from the main office to the project site. To construct the distance we use the longitudinal and latitudinal coordinates of the county of site and the corresponding coordinates for the zip code of the main office. We expect the distance to have some effect on the bids through the cost because, all else equal, it is cheaper to work on projects that are closer to the main office than further so bidders located closer to the site bid low. The correlation between bids and distance is 0.012, which suggests that there are other, more important, factors affecting bids.

One such factor is the backlog, Jofre-Bonet and Pesendorfer [2003], which is defined as the sum of the dollar values of Caltrans projects won but not yet completed by the bidder. Given a backlog, bidders are assumed to work at a constant rate on each of the projects won. But the effect of backlog

on (future) bids depends on the total capacity of the bidder, which we define as the maximum backlog carried by a bidder during the sample period. In other words, following Jofre-Bonet and Pesendorfer [2003], we can use utilization rate defined as the backlog to capacity ratio for each bidder to determine the effect on bids at every auction; see Table A-1.

Bidding Ring. Since our objective is to identify bidders who could be colluding, we need to probe further. We hypothesize that larger projects are more vulnerable to collusive attacks than smaller projects. We note that auctions that are large, valued between $ 1 million and $20 millions attract fewer bidders and most of them tend to be the main bidders. From now on, we focus only on such auctions, and there are a total of 724 projects which total to $2.408 billions (31% of the total), with 413 bidders out of which 202 win at least once. From the summary statistics in Table A-1 we can conclude that: (i) on an average there are slightly more than four bidders; (ii) average winning bid is $3.33 millions, which is less than the average engineers' estimate of $3.77 millions; (iii) the average bid is $3.79 millions; (iv) money on the table – the difference between the highest and the second highest bid – is on average $0.3 million suggesting informational asymmetry among bidders; (v) the distance between the bidder's office and the project site has minimal effect on bids; (vi) on average projects take slightly more than 6 months to be completed; (vii) fringe (type 0) bidders win on average smaller protects than main (type 1) bidders; (viii) on average there are more type 0 bidders participating than type 1 bidders; (ix) the two types differ in their backlogs, and utilization rate which for type0 is double than for type 1.

Moving forward, our working hypothesis is that if there is any bidder who makes a case for a colluder (we define the criteria shortly below) then it must be a type 1 bidder. In Table A-2 we present the descriptive statistics for type 1 bidders who are named A, B, C, etc. (column (i)). To get a sense of market power, we compare the actual number of wins (column (ii)) for each bidder with the bidder's "expected number of wins," (column (iii)) so bidders with consistently higher actual win than the expected wins will be said to have higher market power. Consider bidder A, who bids on a total of 50 projects, $(n_\ell - 1)$ bidders in ℓ^{th} auction, the expected number of wins

is then defined to be $\sum_{\ell=1}^{50} 1/n_\ell$. With the exception of five bidders (A, C, I, L and U), all bidders win more contracts than expected. Column (v) reports the average bid (in millions of $) of each bidder across all participated auction, column (vi) reports the revenue share, which is equal to the total value of the bidder's winning bid as a fraction of the total value of winning bids for all projects; and column (vii) reports the the participation (or bid frequency) rate. Bidder D stands out, with participation rate of 0.44.

Next, we follow Bajari and Ye [2003] to detect bidders whose bidding pattern systemically violate exchangeability and independent bidding, both of which are implied by competitive bidding. To increase the likelihood of picking a coalition we give more emphasis to bidders who participate in the same auction because the literature on collusion suggests that ring members tend to participate in the same auctions to enforce the bidding agreement; see Marshall and Marx [2012]. Like Aryal and Gabrielli [2013] we consider all combinations of pairs and select those bidders that have at least fifteen simultaneous bids. We want as many simultaneous bidding as possible to be very strict on whom we choose as colluders, but in view of the data, we stop at fifteen because beyond that we do not have enough observations left for the test. See Table A-3. There are fourteen pairs involving fourteen bidders who bid frequently together.

To test independence we consider the fourteen pairs of bidders bidding frequently and estimate the following models for fourteen type 1 bidders and the remaining bidders, respectively,

$$BID_{i\ell}/EE_\ell = \gamma_0 + \gamma_1 LDIST_{i\ell} + \gamma_2 CAP_{i\ell} + \gamma_3 UTIL_{i\ell} + \gamma_4 LMDIST_{i\ell} + u_{i\ell} \quad (3)$$

$$BID_{i\ell}/EE_\ell = \alpha_0 + \alpha_1 LDIST_{i\ell} + \alpha_2 CAP_{i\ell} + \alpha_3 UTIL_{i\ell} + \alpha_4 LMDIST_{i\ell} + \varsigma_{i\ell}. \quad (4)$$

Here $\mathrm{LDIST}_{i\ell}$ is the logarithm of distance, $\mathrm{LMDIST}_{i\ell}$ is the logarithm of the distance of bidder i's closest competitor, in the project ℓ, and $UTIL_{i\ell}$ is the utilization rate of the capacity.

We estimate (3) with bidder specific coefficients for the bidders who participate frequently and simultaneously and are listed in Table A-3, and for the rest we use (4). For the estimation, we pool both the equations and include project fixed effects. In total, we estimate 16 different models, where each model includes 723 project dummies and 13 bidder dummies, besides

the afore mentioned regressors.[5] We calculate the correlation between estimated residuals for every pair $\{i, j\}$, i.e., $\rho_{ij} = corr(\hat{u}_{i\ell}, \hat{u}_{j\ell})$, from the fourteen type 1 bidders, and use the Pearson's correlation test for (conditional) independence. Results are presented in Table A-4. We find that except for bidders D and T, we reject the null hypothesis of independence at 5% level.

Next, we test exchangeability among bidders, i.e.,

$$
\begin{aligned}
H_0 &: (\forall i, j, i \neq j), (\forall s \in \{1,2,3,4\}) \qquad \gamma_{is} = \gamma_{js},\\
H_A &: (\exists i, j, i \neq j), (\exists s \in \{1,2,3,4\}) \qquad \gamma_{is} \neq \gamma_{js},
\end{aligned}
$$

at both market level by pooling the fourteen bidders in one group and on a pairwise basis (i.e considering that all fourteen bidders are potential colluders). That is, we conduct two exercises. In the first one we test if a cartel formed by the 14 candidates fails the exchangeability test. Second we conduct the test on a pairwise basis, taking into account all the pairs with at least fifteen simultaneous bids.

Let $T = 3,347$ be the number of observations, m the number of regressors and k the number of constraints implied by H_0. Then under the null the test statistic

$$F = \frac{(SSR_C - SSR_U)/r}{SSR_U/(T-m)} \Rightarrow^d F(r, T-m).$$

At the market level, exchangeability hypothesis imposes that the effect of the four explanatory variables is the same for both potential ring ("the big cartel with 14 bidders") members and the remaining bidders. Since there are thirteen dummies (indexing the bidders) and for each case there are four restrictions (under the null), the total number of restrictions imposed under the null is $r = 52$. Here $m = 747$ and $T - m = 2600$ and the estimated $F-$statistic is 6.2019 with the upper tail area equal to 0. Therefore we reject the null of exchangeability when comparing the fourteen bidders (potential cartel members) against the remaining bidders.

For our exercise to be meaningful, it must be the case that there is no other collusive strategy being used to sustain collusion. In particular, we have implicitly ruled out the possibility that bidders use bid-rotations to sustain

[5] Presenting all the results would require considerable space, and hence we do not present the results, but they are available to the reader upon request.

collusion. Even though a definitive proof that there is no bid rotation is hard to come by, under the assumption that all 14 bidders are members of one ring, we determined all auctions where at least one of these 14 bidders were present. Then for all such auction we visually verified that there is no obvious and discernible pattern in participation. In Figure 1, on the x-axis we have the date for the auction and on the y-axis we put the bidders. Each color corresponds to one unique bidder. And as can be verified there is at least no (visual) evidence of any rotating pattern. This won't work if bidders use some other cunning ways (sun-spot) to coordinate, like in Asker [2008].

It is also known that sustaining large cartel is difficult. We view 14 as being a large cartel. So, to get a tighter number, we impose further conditions that bids must satisfy to be termed as collusive bids, and hope to reduce the size of the ring. We performed the pairwise tests by pooling bidders accordingly and find that the hypothesis of exchangeability is rejected at conventional levels for 12 out of 14 pairs including the pair (D,M), (A,D) and (D,E). See Table A-5 for details.[6] Comparing the "expected win" with the actual win for these pairs, we do see that at least one member of the pair wins often. Comparing Table A-2 and Table A-3 we can conclude that: (i) firm A exclusively bids against firm D; (ii) firm E bids remarkably frequently with both firm A and firm D; (iii) the pairs (D,M) and (A,D) have the highest number of simultaneous bids. All of these suggest that bidders (A,D,E, M) could be considered as potential collusive ring. Based on the previous analysis all pairs of bidders considered do not pass at least one of the tests for competitive bidding. However, taking into account the number of simultaneous bids, bidders D and M bid simultaneously more often than others. And since the triplet (A,D,E) also fit the collusive behavior, we consider colluding bidders to be (A, D, E, M) in our second exercise.

Summary. Using the methods in Bajari and Ye [2003] we tested if bids by type 1 bidders across auctions satisfied independence and exchangeability or not. We found that 14 bidders fail at least one the two criteria. In the counterfactual exercise, we shall treat them as the members of one bidding ring.

[6] Ideally we would have liked to conduct the tests for every subsets of these fifteen bidders not just the pairs, but we have insufficient data to analyze each subset.

The cost thus estimated will provide a (worst-case) upper bound. Recognizing that the difficulty of sustaining collusion increases with the (potential) members, we sought to impose further restrictions on bidding pattern in a hope to minimize the size. At the end we conclude that the smallest collusive ring would have 4 bidders (they do not have to be members of one cartel, but if they are then it again provides an upper bound on the counterfactual cost). In contrast, if we were using a data where it was known that bidders colluded and the identity of the bidders were known, this exercise would have not have been necessary.

5. Estimation Results and Discussion

Here, we take the estimates of the pseudo cost recovered earlier, along with the set of "colluders" (these are those who fail Bajari and Ye [2003] tests and frequently bid in pairs) from above and implement the CM-auction. We ask, how much extra would it cost for Caltrans to procure the same projects using the new auction. Since we have cost for each bid, we can easily compute the change in cost.

Let W_ℓ be the set of colluders who are present in auction ℓ and let $i_\ell \in N_\ell$ be the winner and $\tilde{b}_\ell$ the corresponding bid. Let o_ℓ be the smallest cost amongst all bidders participating in auction ℓ who belong to the coalition W_ℓ except for the winner, ie. o_ℓ is the CM-price. That is if the winner belongs to the coalition, i.e. $i_\ell \in W_\ell$, then $o_\ell = \min_{j \neq i \in n_\ell \cap \{W_\ell \setminus i_\ell\}} \hat{C}_j$ and if the winner is not a member of the coalition then we can set $o_\ell = \tilde{c}_\ell$, the real winning cost. Then the difference between CM-auction and the data is $r_\ell = o_\ell - \tilde{c}_\ell$ if the winner is a cartel member and $r_\ell = 0$ otherwise. Once we compute the change in cost r_ℓ for all auctions the total change in cost of procuring is just $\sum_\ell r_\ell$. For the first case the total set of colluders is the 14 bidders and $W_\ell = W \cap n_{1\ell}$, while for the second case $W_\ell = n_{1\ell} \cap \{A, D, E, M\}$.

When we consider the bidding ring with 14 members, we find that the cost increases by 5%, and it provides the upper bound on the change in cost because larger the coalition larger will be the mark-up. Likewise when we consider the bidding ring with only 4 members, we find that the cost increases by 1.6%, which can be considered as the lower bound. For any other ring sizes the induced increase in cost will be in the range $[1.6\%, 5\%]$.

This exercise, however, ignores the fact that these expenses are covered through (distortionary) taxes. So we have to augment the cost by adding the marginal excess burden (MEB) per additional dollar of tax revenue. An estimate of MEB is hard to come by, but one of the most reliable estimate is by Ballard, Shoven, and Whalley [1985], and they find it to be 0.3. In other words, each dollar raised through taxes leads to a distortion of 30%. Therefore, the total cost of implementing C-M is between 2.08% to 6.5%.

Discussion. The natural question is if this cost is big. Although this is very hard to tell, we believe that this is not a huge cost, especially when factor in that in reality we don't even know how much we lose if bidders are colluding and nothing is done. In some sense, this extra cost is the price of collusion. The real cost of implementing this auction is not the dollar amount but the fact that to run this auction successfully the auctioneer (in particular the public sector that runs the auction) must be willing not to prosecute the colluders. From the legal point of view this requires extensions of the Leniency Program run by the antitrust division of department of Justice. We have nothing to contribute to that strand of literature, but our hope is to make a case that faced with the challenge we out to think outside of the box and explore other options beyond just seeking criminal charges. This is also an example where economics and jurisprudence are both equally important.

One caveat of C-M auction is that it is a static auction and thus ignores any dynamic incentives.[7] Could it be possible that running C-M auction repeatedly will encourage not discourage collusion? C-M auction, however, need not even work in a dynamic bidding environment. We hope to draw the attention of researchers in this area to this lacuna problem in bridging the gap between theoretical mechanism design and its empirical application. In the future, there is a possibility to address this problem by combining the insights from Jofre-Bonet and Pesendorfer [2003] (dynamic auctions) and Che and Kim [2009] (collusion-proof auction with endogenous entry).

[7] See Aubert, Rey, and Kovacic [2006] for the effect of leniency program on cartels.

APPENDIX

A-1. Estimation

In this section we outline the estimation problem and discuss the choice of bandwidths and kernels. To account for the skewness in the bid distribution, a widely observed problem encountered with auction data, we use logarithmic transformation. For notational simplicity we suppress the dependance of the distributions on (X, N), unless otherwise noted. Log transformation of (2) gives

$$c_{kM} = \xi_k(d_k, n) = e^{d_k} - \frac{e^{d_k}}{(n_k - 1)\frac{g_{kd}(d_k|\cdot,\cdot)}{1-G_{kd}(d_k|\cdot,\cdot)} + n_1 \frac{g_{1d}(d_k|\cdot,\cdot)}{1-G_{1d}(d_k|\cdot,\cdot)}} \tag{5}$$

where $d_k = \ln(b_k)$ and $G_{kd}(\cdot|\cdot,\cdot), g_{kd}(\cdot|\cdot,\cdot)$ are the distribution and density of $\log(b_k)$ for type k. Define $K_H(u) = |H|^{-1}K(H^{-1}u)$, where H is a non-singular $d \times d$ matrix, the bandwidth matrix that usually takes the form $H = hI_d$ and $|B|$ denotes its determinant. The observations are given by $\{(Z_i^T, Y_i) : i = 1, \ldots, n\}$ with $Z_i = (X_i, N_{0i}, N_{1i})^T$. Let (x, n_0, n_1) be a point in $\mathbb{R}^3$. The estimators involved are, as mentioned above Local Polynomial Estimators. For our application, $R = 2$ and therefore we implement a LPE(2) for each cdf involved and a LPE (1) for each pdf involved. Let $Y_{p\ell}^G = \mathbb{1}(B_{pl} \leq b)$. Using a local quadratic approximation to estimate each cdf implies obtaining the solution to the following least squares minimization problem

$$\begin{aligned}
&\sum_{\{\ell : I_\ell = i\}}^{L} \sum_{p=1}^{i} \Big\{ Y_{p\ell}^G - \Big[\beta_0 + \beta_1(X_{p\ell} - x) + \beta_2(N_{1\ell} - n_1) + \beta_3(N_{0\ell} - n_0) \\
&+\beta_{11}(X_{p\ell} - x)^2 + \beta_{12}(X_{p\ell} - x)(N_{1\ell} - n_1) + \beta_{13}(X_{p\ell} - x)(N_{0\ell} - n0) \\
&+\beta_{23}(N_{0\ell} - n_0)(N_{1\ell} - n1) + \beta_{22}(N_{1\ell} - n_1)^2 + \beta_{33}(N_{0\ell} - n_0)^2\Big]\Big\}^2 K_H(Z - z)
\end{aligned}$$

with respect to $\beta_G = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{23}, \beta_{22}, \beta_{33})$. In particular we are interested in $\beta_0 = G(b|x, n_0, n_1)$. Then, we know from the least squares theory that $\hat{\beta}_G = (Z_G^T W_G Z_G)^{-1} Z_G^T T_G Y$, where the design matrix Z_G for the local quadratic case (what we use) is

$$Z_G = \begin{pmatrix} 1 & (X_{1,1}-x) & (N_{0,1}-n_0) & (N_{1,1}-n_1) & (X_{1,1}-x)^2 & (X_{1,1}-x)(N_{0,1}-n_0) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (X_{1,n_i}-x) & (N_{0,n_i}-n_0) & (N_{1,n_i}-n_1) & (X_{1,n_i}-x)^2 & (X_{1,n_i}-x)(N_{0,n_i}-n_0) \end{pmatrix.$$

$$\begin{matrix} (X_{1,1}-x)(N_{1,1}-n_1) & (N_{0,1}-n_0)(N_{1,1}-n_1) & (N_{0,1}-n_0)^2 & (N_{1,1}-n_1)^2 \\ \vdots & \vdots & \vdots & \vdots \\ (X_{1,n_i}-x)(N_{1,n_i}-n_1) & (N_{0,n_i}-n_0)(N_{1,n_i}-n_1) & (N_{0,n_i}-n_0)^2 & (N_{1,n_i}-n_1)^2 \end{matrix}\right)$$

For the densities involved define $Y^g_{p\ell} = \frac{1}{h_{2g}} K_{2g}\left(\frac{B_{p\ell}-b}{h_{2g}}\right)$. We use a local linear estimator, i.e. LPE(1) which, as before, is obtained as the solution to the following least squares problem

$$\sum_{\{\ell:I_\ell=i\}}^{L} \sum_{p=1}^{i} \left\{ Y^g_{p\ell} - \beta_0 + \beta_1(X_{p\ell}-x) + \beta_2(N_{1\ell}-n1) + \beta_3(N_{0\ell}-n0) \right\}^2 K_H(Z-z)$$

It is well known that $\hat{\beta}_g = (Z^T T_g Z)^{-1} Z^T T_g Y$. The design matrix Z for the local linear case is

$$Z = \begin{pmatrix} 1 & (X_{1,1}-x) & (N_{0,1}-n_0) & (N_{1,1}-n_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (X_{1,n_i}-x) & (N_{0,n_i}-n_0) & (N_{1,n_i}-x) \end{pmatrix}.$$

The corresponding weighting matrix for each estimation procedure are $T_G = \operatorname{diag}\{K_H(Z_i - z)\}$ and $T_g = \operatorname{diag}\{K_H(Z_i - z)\}$, respectively. The bandwidths and kernels involved for distributions and densities are different.

Choices of Kernels and Bandwidths. Since the exact choice of the Kernels is not crucial for inference, we use product of univariate kernels to represent the multivariate kernel, i.e.

$$K_m\left(\frac{a-A_k}{h_g}, \frac{b-B_k}{h_g}, \frac{n-N_k}{h_{gn}}\right) = K_a\left(\frac{a-A_k}{h_g}\right) K_b\left(\frac{b-B_k}{h_g}\right) K_n\left(\frac{n-N_k}{h_{gn}}\right).$$

Here, $K_m(\cdot,\cdot,\cdot)$ is the multivariate Kernel, $K_a(\cdot)$ and $K_b(\cdot)$ denote the univariate Kernels corresponding to the continuous variables A and B, respectively, and $K_n(\cdot)$ is the kernel for the discrete variables such that $K_n(\cdot) :=$

$K_{n_0}(\cdot) \cdot K_{n_1}(\cdot) \cdot K_{n_2}(\cdot)$. The kernels for continuous variables should be symmetric with bounded supports, so we decided to use the Epanechnikov Kernel function $K(u) = 3/4(1-u^2)\mathbb{1}(|u| \leq 1)$, as it is an optimal Kernel in the sense that it minimizes the asymptotic mean squared error over all non-negative functions Fan, Gasser, Gijbels, Brockmann, and Engel [1993]. For the discrete variables, we use Gaussian Kernel because, as there is less variation in the number of bidders it is desirable to give less weight to observations farther from the point at which estimation takes place and is best achieved with a kernel with unbounded support.[8] We assume the smoothness parameter $R = 2$ for the cost distribution. To ensure uniform consistency at the optimal rates, the bandwidths for the continuous variables are chosen to be $h_g = 1.06 \times 2.214 \times \hat{\sigma} \times (T)^{-1/(2R+1)}$, $h_G = 1.06 \times 2.214 \times \hat{\sigma} \times (T)^{-1/(2R+3)}$. The constant term comes from the so–called rule of thumb and the factor 2.214 is the one corresponding to the use of Epanechnikov Kernels instead of Gaussian Kernels; see Härdle [1991].

A-2. Tables and Figures

TABLE A-1. Summary Statistics

	Observations			Mean			Std. Dev.			Minimum			Maximum		
	All	Type 0	Type 1	All	Type 0	Type 1	All	Type 0	Type 1	All	Type 0	Type 1	All	Type 0	Type 1
# Bidders	724	413	311	4.62	4.97	4.16	2.37	2.55	2.01	2	2	2	23	23	13
Winning bid	724	413	311	3.33	2.82	4.00	3.11	2.74	3.42	0.43	0.43	0.52	19.88	19.19	19.88
Money	724	413	311	0.30	0.25	0.37	0.46	0.39	0.53	0.00	0.00	0.00	4.52	3.72	4.52
Eng. Est.	724	413	311	3.77	3.23	4.48	3.49	3.12	3.83	1.00	1.00	1.00	19.96	19.96	19.72
All bids	3347	2274	1073	3.79	3.57	4.24	3.51	3.40	3.69	0.43	0.43	0.52	24.01	24.01	23.11
Backlog	3347	2274	1073	4.30	0.85	11.60	9.76	2.42	14.37	0.00	0.00	0.00	61.80	26.53	61.80
Distance	3347	2274	1073	123.98	119	134	162.93	175	134	1.32	1	3	1578.85	1579	897
Capacity	413	392	21	2.30	1.41	18.88	5.69	3.50	11.11	0	0	7.23	61.80	26.53	61.80
Utilization	3347	2274	1073	0.20	0.15	0.32	0.32	0.31	0.31	0	0	0	1	1	1

Note: This table presents the summary statistics for both types. Winning bids is the lowest bid at which the auction is awarded. Money, stands for money-on-the-table and is equal to the different between second lowest bid and the lowest bid. Eng. Est. stands for the engineer's estimate of the level of expenditure. $\text{Util}_{\text{it}} = \text{Backlog}_{\text{it}}/\text{Cap}_{\text{i}}$ (if $Cap_i = 0$, then $Util_{it} = 0$ for all t)

[8] There are no theoretical restrictions to the kernels applied to discrete variables.

TABLE A-2. Type 1 Bidders

Bidder ID (i)	# of Bids (ii)	# of wins (iii)	Exp. # of wins (iv)	Avg. bid (in ml. $) (v)	Revenue Share (vi)	Participation rate (vii)
A	50	9	10.34	4.83	0.02	0.07
B	34	13	10.51	3.21	0.01	0.05
C	43	9	10.46	5.32	0.01	0.06
D	319	97	87.32	3.61	0.14	0.44
E	46	11	10.15	4.49	0.02	0.06
F	42	15	10.70	3.63	0.02	0.06
G	25	12	5.84	4.09	0.03	0.03
H	26	6	5.16	5.03	0.01	0.04
I	34	4	6.90	8.44	0.02	0.05
J	35	16	7.95	4.32	0.02	0.05
K	29	13	6.94	3.69	0.02	0.04
L	31	5	6.82	6.37	0.01	0.04
M	50	16	12.95	4.03	0.03	0.07
N	33	9	6.31	3.35	0.02	0.05
O	28	10	8.10	3.48	0.01	0.04
P	47	12	8.82	4.37	0.02	0.06
Q	25	13	5.99	3.75	0.02	0.03
R	68	16	15.22	4.77	0.03	0.09
S	26	7	4.78	5.75	0.03	0.04
T	41	11	7.18	2.92	0.02	0.06
U	41	7	10.27	4.50	0.02	0.06
Total	1073	311	259		0.52	

Notes: Type 1 are the bidders with revenue shares of at least 1% and participation rate of at least 3%. Column: (i) is the bidder identity; (ii) is the number of (total) bids across all auctions; (iii) is the actual number of wins; (iv) is the expected number of wins; (v) the average bid (in million $) across all auctions; (vi) is the revenue share defined as the ratio of the sum of all winning bids by the bidder to the total of all winning bids; and (vii) is the participation rate is the fraction of the time the bidder participates.

TABLE A-3. Summary of Simultaneous Bids

Pairs (i)	# of Smlt. Bids (ii)	Exp. # of Wins (iii)	First Wins (iv)	Second Wins (v)
(A,D)	44	9.03	9	5
(A,E)	20	4.05	3	6
(B,D)	29	9.51	12	10
(C,D)	17	5.65	5	9
(D,E)	41	8.67	8	9
(D,F)	26	7.46	5	9
(D,H)	19	3.92	7	3
(D,L)	25	5.16	7	5
(D,M)	44	11.08	13	14
(D,O)	27	7.96	10	10
(D,S)	22	4.20	5	6
(D,T)	19	2.97	2	3
(K,U)	22	4.91	11	2
(T,U)	15	2.81	5	2

Notes: Summary of simultaneous bids. Column: (i) is the pair of bidders; (ii) is the number of simultaneous bids, i.e, the number of auctions where the pair from column (i) participate; (iii) is the expected number of wins; (iv) is the number of auctions where the first bidder in the pair wins; (v) is the number of auctions where the second bidder from the pair wins.

TABLE A-4. Conditional Independence Test

Pairs (i)	F- Stat (ii)	p-value (iii)	n (iv)	d.o.f (v)
(A,D)	0.7660	0.0000	44	42
(A,E)	0.7427	0.0002	20	18
(B,D)	0.7331	0.0000	29	27
(C,D)	0.9239	0.0000	17	15
(D,E)	0.6530	0.0000	41	39
(D,F)	0.7570	0.0000	26	24
(D,H)	0.4734	0.0406	19	17
(D,L)	0.7643	0.0000	25	23
(D,M)	0.8538	0.0000	44	42
(D,O)	0.8555	0.0000	27	25
(D,S)	0.6877	0.0004	22	20
(D,T)	0.4305	0.0658	19	17
(K,U)	0.6529	0.0010	22	20
(T,U)	0.6271	0.0123	15	13

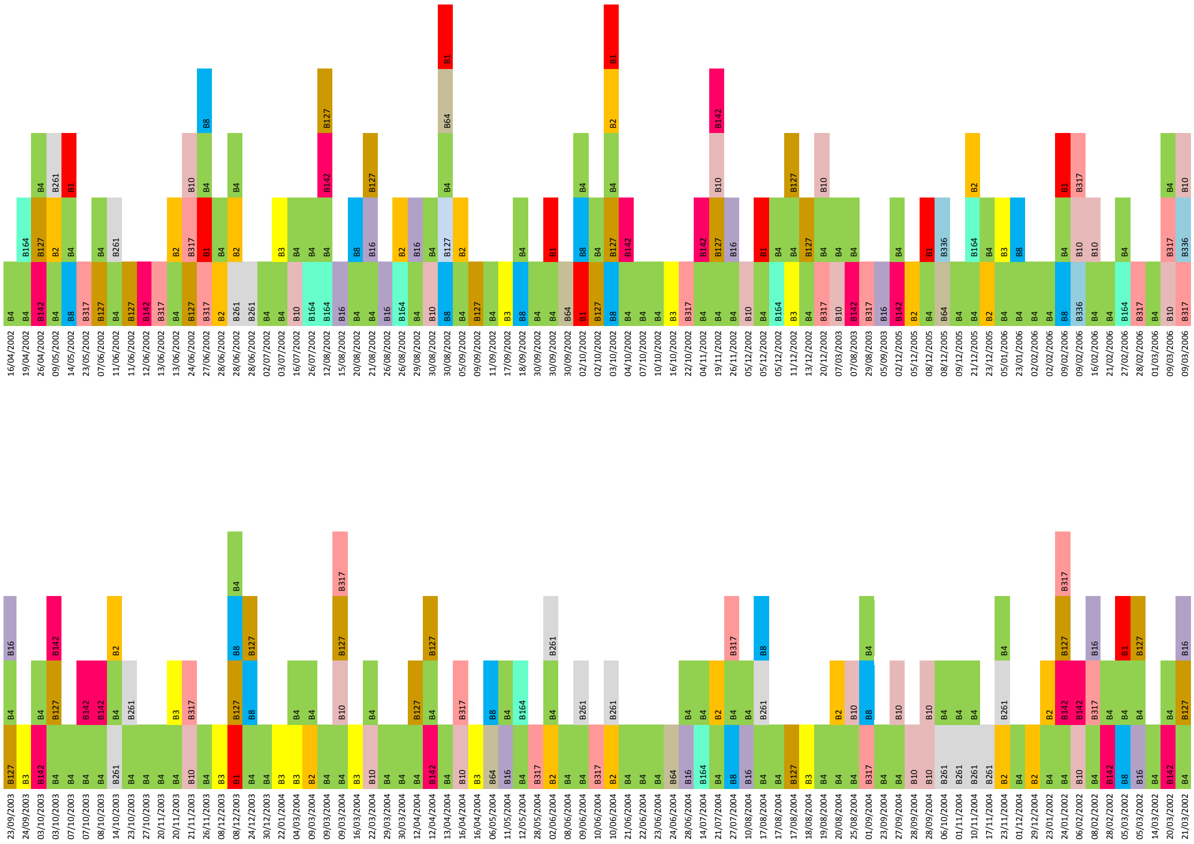
Notes: This table provides result from conditional independent test for each pair. Column: (i) identifies the pair; (ii) is the F-statistic; (iii) is the corresponding p-value; (iv) is the number of observation used in the test; (v) is the degrees of freedom.

TABLE A-5. Exchangeability Test on Pairwise Basis

Pairs (i)	F -stat (ii)	UTA (iii)
(A,D)	5.1926	0.0000
(A,E)	2.3506	0.0162
(B,D)	5.9269	0.0000
(C,D)	8.4150	0.0000
(D,E)	7.9435	0.0000
(D,F)	6.8343	0.0000
(D,H)	5.1926	0.0000
(D,L)	5.5009	0.0000
(D,M)	7.1579	0.0000
(D,O)	6.1059	0.0000
(D,S)	3.7330	0.0002
(D,T)	4.6150	0.0000
(K,U)	0.7972	0.6050
(T,U)	0.3504	0.9460

Notes: This table provides result from exchangeability test for each pair. Column: (i) identifies the pair; (ii) is the F-statistic; (iii) is the upper tail area (UTA).

FIGURE 1. No Bid Rotation



Notes: This figure shows that there is no bid rotation pattern in the data. On the X-axis is the date of all auctions where at least one of type 1 bidder is present. Each such bidder is the given a unique identifier, and a unique color for visual aid. Consider the first auction on 16-04-2002 where only bidder B4 participates. The next auction is on 19-06-2002 where B4 and another bidder B164 participates, and so on.

REFERENCES

ABADIE, A., AND G. W. IMBENS (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267. 8

ARYAL, G., AND M. F. GABRIELLI (2013): "Testing for Collusion in Asymmetric First Price Auction," *International Journal of Industrial Organization*, 31, 26–35. 2, 13

ARYAL, G., M. F. GABRIELLI, AND Q. VUONG (2015): "Semiparametric Estimation of First-Price Auction Models," *http://arxiv.org/abs/1407.7140*. 3, 6, 8

ASKER, J. (2008): "A Study of the Internal Organisation of a Bidding Cartel," *American Economic Reveiw*, 100(3), 724–762. 2, 15

ATHEY, S., AND P. HAILE (2006): "Empirical Models of Auctions," *Cowles Foundation Discussion Paper*, 1562. 1

AUBERT, C., P. REY, AND W. E. KOVACIC (2006): "The Impact of Leniency and Whistle-Blowing Programs on Cartels," *International Journal of Industrial Organization*, 24(6), 1241–1266. 17

AUSUBEL, L. M., AND P. MILGROM (2006): "The Lovely but Lonely Vickrey Auction," in *Combinatorial Auctions*, ed. by P. Cramton, Y. Shoham, and T. Steinberg, pp. 17–40. MIT Press. 1

BAJARI, P. (2001): "Comparing Competition and Collusion: A Numerical Approach," *Economic Theory*, 18, 187–205. 2

BAJARI, P., AND L. YE (2003): "Deciding Between Competition and Collusion," *Review of Economics and Statistics*, 85, 971–989. 2, 3, 10, 13, 15, 16

BALDWIN, L., R. MARSHALL, AND J. RICHARD (1997): "Bidder Collusion at Forest Service Timber Sales," *Journal of Political Economy*, 4, 657–699. 2

BALLARD, C. L., J. B. SHOVEN, AND J. WHALLEY (1985): "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States," *American Economic Reveiw*, 75(1), 128–138. 4, 17

BIERENS, H. (1987): "Kernel Estimators of Regression Functions," in *Advances in Econometrics: Fifth World Congress*, ed. by T. Bewley, vol. 1, chap. 3, pp. 99–144. Cambridge University Press. 8

BORRELL, J.-R., J. L. JIMÉNEZ, AND C. GARCIA (2013): "Evaluating Antitrust Leniency Programs," *Journal of Competition Law and Economics*, 10(1), 107–136. 4

CHE, Y.-K., AND J. KIM (2009): "Optimal Collusion-Proof Auctions," *Journal of Economic Theory*, 144, 565–603. 17

CHEN, J., AND S. MICALI (2012): "Collusive Dominant-Strategy Truthfulness," *Journal of Economic Theory*, 147(3), 1300–1312. 1, 2, 4

CLINE, D., AND J. D. HART (1991): "Kernel Estimation of Densities of Discontinuous Derivatives," *Statistics*, 22, 69–84. 8

COMANOR, W. S., AND M. A. SCHANKERMAN (1976): "Identical Bids and Cartel Behavior," *Bell Journal of Economics*, 7, 281–286. 2

CONLEY, T. G., AND F. DECAROLIS (forthcoming): "Detecting Bidders Groups in Collusive Auctions," *AEJ: Microeconomics*. 2

FAN, J., T. GASSER, I. GIJBELS, M. BROCKMANN, AND J. ENGEL (1993): "Local Polynomial Fitting: A standard for Nonparametric Regression," *Mimeo, Series 2302, Institute of Statistics, UNC- Chapel Hill*. 20

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall. 3, 7

FEINSTEIN, J. S., M. K. BLOCK, AND F. C. NOLD (1985): "Asymmetric Information and Collusive Behavior in Auction Markets," *The American Economic Review*, 75(3), 441–460. 2

GASSER, T., AND H. G. MÜLLER (1979): "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation*, ed. by T. Gasser, and M. Rosenblatt, vol. 757 of *Lecture Notes in Mathematics*. Springer-Verlag Hidelberg. 8

GASSER, T., H. G. MÜLLER, AND V. MAMMITZSCH (1985): "Kernels for Nonparametric Curve Estimation," *Journal of Royal Statistical Society. Series B (Methodological)*, 47, 238–252. 8

GREEN, J., AND J.-J. LAFFONT (1979): "On Coalition Incentive Compatibility," *The Review of Economic Studies*, 46(2), 2430254. 2

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574. 3, 6, 8

HÄRDLE, W. (1991): *Smoothing Techniques with Implementation in S*. Springer-Verlag New York, Inc. 20

HARRINGTON, J. E. (2008): "Detecting Cartels," in *Handbook in Antitrust Economics*, ed. by P. Buccirossi. MIT Press. 2

HOU, D., AND D. SKEIE (2013): "LIBOR: origins, economics, crisis, scandal and reform," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan. 2

JOFRE-BONET, M., AND M. PESENDORFER (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 1443–1489. 11, 12, 17

JONES, M. (1993): "Simple Boundary Correction for Kernel Density Estimation," *Statistics and Computing*, 3, 135–146. 8

LAFFONT, J.-J., AND D. MARTIMORT (1997): "Collusion under Asymmetric Information," *Econometrica*, 65, 875–911. 2

——— (1998): "Collusion and Delegation," *RAND Journal of Ecoomics*, 29, 280–305. 2

——— (2000): "Mechanism Design with Collusion and Correlation," *Econometrica*, 68(2), 309–342. 2

LANG, K., AND R. W. ROSENTHAL (1991): "The Contractor's Game," *The RAND Journal of Economics*, 22, 329–338. 2

MARSHALL, R., AND M. MEURER (2004): "Bidder Collusion and Antitrust Law: Refining the Analysis of Price Fixing to Account for the Special Features of Auction Markets," *Antitrust Law Journal*, 72(1), 83–118. 2

MARSHALL, R. C., AND L. M. MARX (2009): "The Vulnerability of Auctions to Bidder Collusion," *Quarterly Journal of Economics*, 124(2), 883–910. 1

——— (2012): *The Economics of Collusion: Cartels and Bidding Rings*. MIT Press. 2, 13

PESENDORFER, M. (2000): "A Study of Collusion in First–Price Auctions," *Review of Economic Studies*, 67, 381–411. 2

PORTER, R., AND D. ZONA (1993): "Detection of Bid–Rigging in Procurement Auctions," *Journal of Political Economy*, 101, 518–538. 2

——— (1999): "Ohio School Milk Markets: An Analysis of Bidding," *Rand Journal of Economics*, 30, 263–288. 2

SCHUSTER, E. (1985): "Incorporating Support Constraints Into Nonparametric Estimators of Densities," *Communication in Statistics- Theory and Methods*, 14, 1123–1136. 8

SILVERMAN, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York. 8

STONE, C. (1982): "Optimal Rates of Convergence for Nonparametric Regressions," *Annals of Statistics*, 10, 1040–1053. 8

TAIBBI, M. (2012): *The Scam Wallstreet Learned from the Mafia.*The Roling Stones. 2

——— (2013): *Everything IS Rigged: The Biggest Price-Fixing Scandal Ever.*The Roling Stones. 2

ZHANG, S., AND R. KARUNAMUNI (2000): "On Nonparametric Density Estimation at the Boundary," *Nonparametric Statistics*, 12, 197–221. 8